1 आंकड़े : स्रोत और संकलन

आप आंकड़ों के विभिन्न प्रकार देख और उपयोग कर चुके हैं। उदाहरण के लिए, दूरदर्शन पर प्रत्येक समाचार बुलेटिन के अंत में, मुख्य शहरों के अभिलिखित तापमान प्रदर्शित किये जाते हैं। उसी प्रकार, भारत के भूगोल पर लिखी गई पुस्तकें, जनसंख्या की वृद्धि एवं वितरण और विभिन्न फसलों, खनिजों और औद्योगिक उत्पादों संबंधी आंकड़ों को तालिका के रूप में दर्शाती हैं। क्या आपने कभी सोचा है कि उनका अर्थ क्या है? ये आंकड़े कहाँ से प्राप्त किए जाते हैं? अर्थपूर्ण सूचनाएँ प्राप्त करने के लिए वे किस प्रकार से तालिकाबद्ध एवं प्रक्रमित किए जाते हैं। इस अध्याय में हम आंकड़ों के इन पक्षों पर विचार-विमर्श करेंगे और इन अनेक प्रश्नों के उत्तर देने का प्रयत्न करेंगे।

आंकड़े क्या हैं?

आंकड़ों को ऐसी संख्याओं के रूप में परिभाषित किया गया है जो यथार्थ विश्व के मापन को प्रदर्शित करती हैं। आधार सामग्री एकमात्र माप है। हम प्राय: समाचार पढ़ते हैं, जैसे–बाड़मेर में लगातार 20 से.मी. वर्षा अथवा चौबीस घंटों में बाँसवाड़ा में निरंतर 35 से.मी. वर्षा अथवा सूचना जैसे–रेलगाड़ी द्वारा नयी दिल्ली-मुंबई की दूरी, वाया कोटा-वड़ोदरा 1305 कि.मी. है और वाया इटारसी-मनमाड 1542 कि.मी. है। यह संख्यात्मक सूचना आंकड़ा कहलाती है। यह आसानी से अनुभव किया जा सकता है कि आज के संसार में बड़ी संख्या में आंकड़े उपलब्ध हैं फिर भी इन आंकड़ों से तार्किक निष्कर्ष निकालना उस समय कठिन हो जाता है जबकि ये अपरिष्कृत रूप में होते हैं। इसलिए यह सुनिश्चित कर लेना महत्वपूर्ण है कि मापी गई सूचना प्रतीक गणितीय रूप से प्राप्त की गई है अथवा तार्किक रूप से निगमित किए गए हैं अथवा सांख्यिकीय विधि से परिकलित किए गए हैं। सूचना को एक प्रश्न के अर्थपूर्ण उत्तर अथवा अर्थपूर्ण उद्दीपक के रूप में परिभाषित किया गया है जिसे अगले प्रश्नों में सोपानित किया जा सकता है।

आंकड़ों की आवश्यकता

भौगोलिक अध्ययन में मानचित्र एक महत्वपूर्ण साधन है। इसके अतिरिक्त परिघटनाओं के वितरण और वृद्धि को सारणीबद्ध रूप में आंकड़ों के द्वारा स्पष्ट किया गया है। हम जानते हैं कि पृथ्वी की सतह पर बहुत-सी परिघटनाओं के मध्य अंतर्संबंध होते हैं। ये अन्योन्य क्रियाएँ बहुत से चरों द्वारा प्रभावित होती हैं जिनकी सबसे

अच्छी व्याख्या मात्रात्मक रूप में की जा सकती है। आज उन चरों का सांख्यिकीय विश्लेषण आवश्यक हो गया है। उदाहरण के लिए किसी क्षेत्र के शस्य प्रारूप के अध्ययन के लिए, फसल के अंतर्गत क्षेत्र, फसल की उत्पादकता और उत्पादन, सिंचित क्षेत्र, वर्षा की मात्रा और उर्वरक, कीटनाशक और पीड़कनाशी के प्रयोग जैसे निवेश के बारे में सांख्यिकीय सूचना का होना आवश्यक है। इसी प्रकार से किसी क्षेत्र में एक नगर के विकास के अध्ययन के लिए कुल जनसंख्या, घनत्व, प्रवासियों की संख्या, लोगों के व्यवसाय, उनके वेतन, उद्योगों, यातायात और संचार के साधनों से संबंधित आंकड़े आवश्यक होते हैं। इस प्रकार, आंकड़े भौगोलिक विश्लेषण में एक महत्वपूर्ण भूमिका निभाता है।

आंकड़ों का प्रस्तुतीकरण

आपने एक व्यक्ति की कहानी सुनी होगी जो अपनी पत्नी और पाँच साल के बच्चे के साथ यात्रा कर रहा था। रास्ते में उसे एक नदी पार करनी थी। सबसे पहले उसने चार बिंदुओं की गहराई 0.6, 0.8, 0.9, 1.5 मीटर के रूप में मापी। उसने औसत गहराई 0.95 मीटर निकाली। उसके बच्चे की लंबाई 1 मीटर थी। इसलिए उसने उसे नदी पार करने के लिए उतार दिया और उसका बच्चा नदी में डूब गया। दूसरे किनारे पर वह चिंतन करता हुआ बैठ गया, "लेखा-जोखा थाए, तो बच्चा डूबा काहे?" (बच्चा क्यों डूब गया जब गहराई सभी की पहुँच में थी?) इसे सांख्यिकीय दोष कहते हैं जो कि आपको यथार्थ स्थिति से भ्रमित कर सकता है। इसलिए तथ्यों और आकार को जानने के लिए आंकड़ों को एकत्र करना बहुत आवश्यक है, लेकिन उतना ही महत्वपूर्ण आंकड़ों का प्रस्तुतीकरण है। आज सांख्यिकीय विधियों का उपयोग विश्लेषण, प्रस्तुतीकरण और निष्कर्षों को निकालने में भूगोल सहित लगभग सभी शास्त्रों में जो कि आंकड़ों का उपयोग करते हैं, एक महत्वपूर्ण भूमिका निभाता है। इससे यह अनुमान लगाया जा सकता है कि परिघटनाओं का सांद्रण भिन्न पाया जाता है। जैसे कि जनसंख्या, वन अथवा यातायात या संचार नेटवर्क न केवल स्थान और समय के अनुसार बल्कि आंकड़ों के उपयोग से आसानी से समझाया जा सकता है। अन्य शब्दों में आप कह सकते हैं कि चरों के बीच संबंधों की व्याख्या करने में गुणात्मक विश्लेषण से मात्रात्मक विश्लेषण में स्थानांतरण है। इसलिए इन दिनों विश्लेषणात्मक साधन और तकनीकें, विषय को और अधिक तार्किक बनाने और परिशुद्ध निष्कर्ष प्राप्त करने के लिए अत्यधिक महत्वपूर्ण हो गए हैं। आंकड़ों के एकत्रण और संकलन के आरंभ से ही उनके सारणीयन, संगठन, क्रमबद्धता और संक्रियात्मक विश्लेषण तक जब तक कि निष्कर्ष प्राप्त न हो जाए परिशुद्ध सांख्यिकीय तकनीकों का प्रयोग किया जाता है।

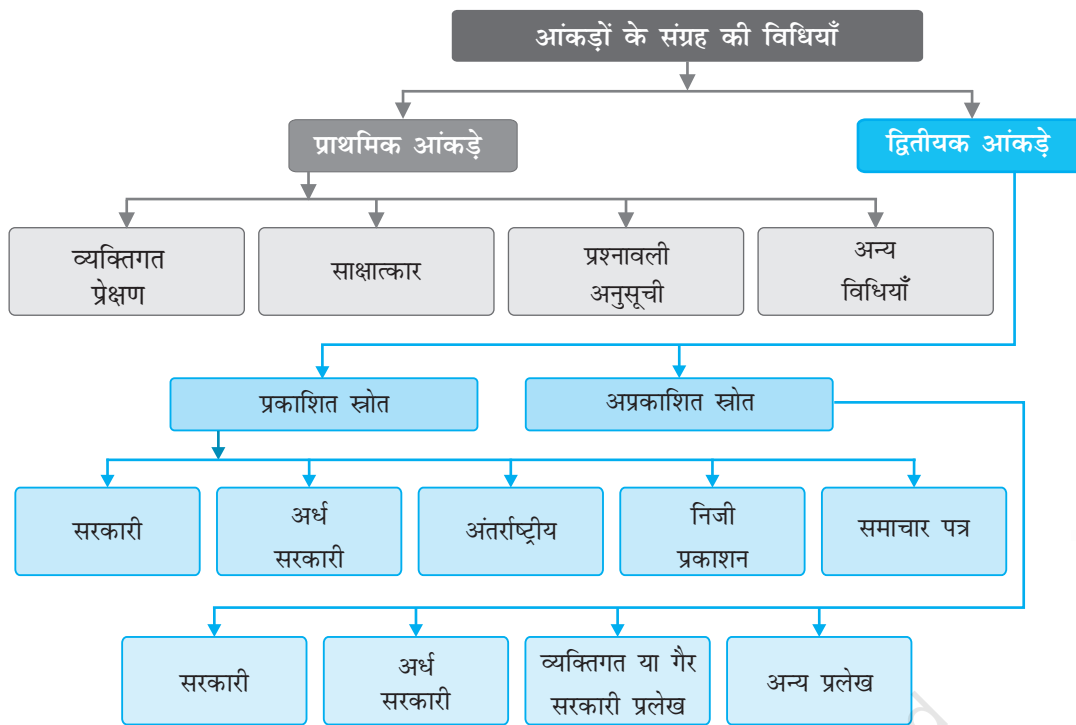
आंकड़ों के स्रोत

आंकड़े निम्नलिखित विधियों से एकत्रित किए जाते हैं–1. प्राथमिक स्रोत 2. द्वितीयक स्रोत।

जो आंकड़े प्रथम बार व्यक्तिगत रूप से अथवा व्यक्तियों के समूह संस्था/संगठन द्वारा एकत्रित किए जाते हैं, आंकड़ों के प्राथमिक स्रोत कहलाते हैं। दूसरी तरफ़ जो आंकड़े किसी प्रकाशित अथवा अप्रकाशित साधनों द्वारा एकत्र किए जाते हैं, द्वितीयक स्रोत कहलाते हैं। *चित्र 1.1* में आंकड़ा संग्रह की विभिन्न विधियाँ दर्शाई गई हैं।

प्राथमिक आंकड़ों के साधन

1. व्यक्तिगत प्रेक्षण

यह सूचनाओं के उस संग्रह की ओर संकेत करता है जो व्यक्तिगत या व्यक्तियों के समूह द्वारा क्षेत्र में प्रत्यक्ष प्रेक्षण द्वारा एकत्र किया जाता है। क्षेत्र-सर्वेक्षण के द्वारा भू-आकृति के लक्षणों, अपवाह प्रारूप, मिट्टी और प्राकृतिक वनस्पति के प्रकारों के साथ-साथ जनसंख्या संरचना, लिंग अनुपात, साक्षरता, परिवहन और संचार के साधन, नगरीय और ग्रामीण अधिवास आदि के बारे में सूचनाएँ एकत्र की जाती हैं। फिर भी व्यक्तिगत

चित्र 1.1 : आंकड़ों के संग्रह की विधियाँ

प्रेक्षण करते समय उसमें सम्मिलित व्यक्ति/व्यक्तियों को निष्पक्ष मूल्यांकन के लिए विषय का सैद्धांतिक ज्ञान और वैज्ञानिक दृष्टिकोण होना चाहिए।

2. साक्षात्कार

इस विधि में शोधकर्ता उत्तर देने वाले से प्रत्यक्ष सूचना संवाद और बातचीत द्वारा प्राप्त करता है। फिर भी, साक्षात्कारकर्ता को क्षेत्र के लोगों से साक्षात्कार करते समय निम्नलिखित सावधानियों को बरतना चाहिए–

(i) लोगों से साक्षात्कार द्वारा जिन सूचनाओं को इकट्ठा करना है, उन विषयों की एक परिशुद्ध सूची तैयार कर लेनी चाहिए।

(ii) साक्षात्कार लेने वाले व्यक्ति/व्यक्तियों को सर्वेक्षण के उद्देश्यों के बारे में स्पष्ट जानकारी होनी चाहिए।

(iii) कोई भी संवेदनशील प्रश्न पूछने से पहले, उत्तर देने वालों को विश्वास में लेना चाहिए और उसे यह विश्वास दिलाना चाहिए कि गोपनीयता बनाई रखी जाएगी।

(iv) अनुकूल वातावरण होना चाहिए जिससे उत्तर देने वाला बिना झिझक के तथ्यों को स्पष्ट कर सके।

(v) प्रश्नों की भाषा साधारण और शिष्ट होनी चाहिए जिससे उत्तर देने वाला प्रेरित होकर सहज ही प्रश्नों से संबंधित सूचना देने के लिए सहमत हो जाए।

(vi) ऐसे प्रश्नों को पूछने से बचना चाहिए जिससे उत्तर देने वालों के आत्मसम्मान अथवा धार्मिक भावनाओं को ठेस न पहुँचे।

(vii) साक्षात्कार के अंत में उत्तर देने वालों से पूछना चाहिए कि वह जो सूचना दे चुके हैं, इसके अतिरिक्त और क्या जानकारी दे सकते हैं?

(viii) उन्हें आपके लिए अपना बहुमूल्य समय प्रदान करने के लिए धन्यवाद और कृतज्ञता ज्ञापित करनी चाहिए।

3. प्रश्नावली अनुसूची

इस विधि में, साधारण प्रश्नों और उनके संभावित उत्तर एक सादे कागज़ पर लिखे रहते हैं और उत्तर देने वालों को दिए गए विकल्पों में से सही उत्तर पर निशान लगाना होता है। कई बार प्रश्नावली में संरचनात्मक प्रश्नों का एक समूह लिखा रहता है और उत्तर देने वालों के विचार जानने के लिए पर्याप्त स्थान दिया रहता है। यदि केवल विवृत्तांत प्रश्नों के माध्यम से लोगों के विचारों को एकत्र करने की जरूरत है तो इसे प्रश्नावली कहते हैं। प्रश्नावली में सर्वेक्षण के उद्देश्य स्पष्ट रूप से उल्लिखित होने चाहिए। यह विधि बड़े क्षेत्र के सर्वेक्षण के लिए उपयोगी होती है। प्रश्नावली को दूरवर्ती क्षेत्रों में भी भेजा जा सकता है। इस विधि की सीमा यह है कि आवश्यक सूचनाओं को उपलब्ध कराने के लिए केवल साक्षर और शिक्षित लोगों से ही संपर्क किया जा सकता है। प्रश्नावली से मिलती-जुलती जिसमें जाँच-पड़ताल से जुड़े प्रश्न दिए रहते हैं, उसे अनुसूची कहा जाता है। प्रश्नावली और अनुसूची में केवल यह अंतर होता है कि प्रश्नावली में उत्तर देने वाला प्रश्नावलियों को स्वयं भरता है जबकि सूची में परिगणक उत्तर देने वाले से प्रश्न पूछकर स्वयं भरता है। प्रश्नावली की तुलना में अनुसूची का मुख्य लाभ यह है कि इसके द्वारा सूचना शिक्षित और अशिक्षित दोनों ही उत्तर देने वालों से एकत्र की जा सकती हैं। एक अनुसूची को भरने के लिए गणनाकर्ता को पूरी तरह प्रशिक्षित होना चाहिए।

4. अन्य विधियाँ

मृदा और जल के गुणों से संबंधित आंकड़े सीधे क्षेत्रों से, मृदा किट और जल गुणवत्ता किट का उपयोग करते हुए उनकी विशेषताओं को माप कर एकत्र किए जाते हैं। इसी तरह क्षेत्र-वैज्ञानिक के उपयोग से फसलों और वनस्पति के स्वास्थ्य के बारे में आंकड़े इकट्ठे कर रहे हैं (*चित्र 1.2*)।

चित्र 1.2 : फसल के स्वास्थ्य की माप लेते हुए क्षेत्र वैज्ञानिक

आंकड़ों के द्वितीयक स्रोत

द्वितीयक स्रोतों के अंतर्गत आंकड़ों के प्रकाशित और अप्रकाशित स्रोत आते हैं जिनमें सरकारी प्रकाशन, प्रलेख और रिपोर्टें सम्मिलित किए जाते हैं।

प्रकाशित साधन

1. सरकारी प्रकाशन

विभिन्न मंत्रालयों और भारत सरकार के विभागों, राज्य सरकारों के प्रकाशन और ज़िलों के बुलेटिन द्वितीयक सूचनाओं के महत्वपूर्ण साधन हैं। इनके अंतर्गत भारत के महापंजीयक कार्यालय द्वारा प्रकाशित भारत की जनगणना, राष्ट्रीय प्रतिदर्श सर्वेक्षण की रिपोर्टें, भारतीय मौसम विज्ञान विभाग की मौसम रिपोर्ट, राज्य सरकारों द्वारा प्रकाशित सांख्यिकीय सारांश और विभिन्न आयोगों द्वारा प्रकाशित आवधिक रिपोर्टें सम्मिलित किए जाते हैं। कुछ सरकारी प्रकाशन *चित्र 1.3* में दर्शाए गए हैं।

चित्र 1.3 : कुछ सरकारी प्रकाशन

2. अर्ध सरकारी प्रकाशन

इस श्रेणी के अंतर्गत नगर विकास प्राधिकरणों और विभिन्न नगरों और शहरों के नगर–निगमों और ज़िला परिषदों के प्रकाशन और रिपोर्ट आते हैं।

3. अंतर्राष्ट्रीय प्रकाशन

अंतर्राष्ट्रीय प्रकाशनों के अंतर्गत वार्षिकी, संयुक्त राष्ट्र के विभिन्न अभिकरणों जैसे–संयुक्त राष्ट्र अभिकरण, वैज्ञानिक तथा सांस्कृतिक संगठन (यूनेस्को), संयुक्त राष्ट्र विकास कार्यक्रम (यू. न. डी. पी.), विश्व स्वास्थ्य संगठन (डब्ल्यू. एच. ओ.), खाद्य व कृषि परिषद् (एफ. ए. ओ.) आदि द्वारा प्रकाशित रिपोर्ट और मोनोग्राफ़ सम्मिलित किए जाते हैं। संयुक्त राष्ट्र के कुछ महत्वपूर्ण प्रकाशन जो आवधिक छपते हैं, वे हैं– डैमोग्राफ़िक इयर बुक, स्टेटिस्टीकल इयर बुक और मानव विकास रिपोर्ट *(चित्र 1.4)*।

चित्र 1.4 : कुछ संयुक्त राष्ट्र प्रकाशन

4. निजी प्रकाशन

इस श्रेणी के अंतर्गत समाचारपत्र और निजी संस्थाओं द्वारा प्रकाशित वार्षिकी पुस्तिका, सर्वेक्षण शोध रिपोर्ट और प्रबंध आते हैं।

5. समाचारपत्र और पत्रिकाएँ

दैनिक समाचारपत्र और साप्ताहिक, पाक्षिक और मासिक पत्रिकाएँ द्वितीयक आंकड़ों के आसानी से प्राप्य स्रोत हैं।

6. इलेक्ट्रॉनिक

यह स्रोत वर्तमान में इलेक्ट्रॉनिक माध्यम विशेषकर इंटरनेट, द्वितीयक आंकड़ों का एक महत्वपूर्ण स्रोत बनकर उभरा है।

अप्रकाशित साधन

1. सरकारी प्रलेख

द्वितीयक आंकड़ों के अन्य स्रोत अप्रकाशित रिपोर्टें, मोनोग्राफ़ और प्रलेख हैं। ये प्रलेख सरकार के विभिन्न स्तरों पर अप्रकाशित रिकार्ड के रूप में तैयार किए और अनुरक्षित रखे जाते हैं। उदाहरण के लिए, गाँव के स्तर पर, राजस्व अभिलेख गाँव के पटवारियों के द्वारा बनाए जाते हैं जो एक गाँव स्तर की सूचना का महत्वपूर्ण साधन हैं।

2. अर्ध सरकारी प्रलेख

अर्ध सरकारी प्रलेखों में, विभिन्न नगर निगम, ज़िला परिषदों और लोक सेवा विभागों द्वारा तैयार और अनुरक्षित की गई आवधिक रिपोर्टें और विकास योजनाएँ सम्मिलित की जाती हैं।

3. निजी प्रलेख

इसके अंतर्गत कंपनियों, व्यापार संघों, विभिन्न राजनैतिक और अराजनैतिक संगठनों और निवासीय कल्याण संघों के अप्रकाशित रिपोर्ट और रिकार्ड सम्मिलित किए जाते हैं।

आंकड़ों का सारणीयन और वर्गीकरण

प्राथमिक अथवा द्वितीयक साधनों द्वारा एकत्र किए गए आंकड़े प्रारंभ में बहुत कम समझ में आने वाली सूचनाओं के एक उलझे समूह के रूप में दिखाई देते हैं। यह आंकड़ा संरचना कच्चा आंकड़ा कहलाती है। अर्थपूर्ण निष्कर्ष निकालने और उपयोग में लाने के लिए उन अपरिष्कृत कच्चे आंकड़ों के सारणीयन और वर्गीकरण की जरूरत होती है।

सांख्यिकीय सारणी, आंकड़ों को संक्षिप्त करने और प्रस्तुत करने के सबसे साधारण उपायों में से एक है। यह आंकड़ों की कॉलम और पंक्तियों में की गई एक सुव्यवस्थित व्यवस्था है। इस सारणी का उद्देश्य प्रस्तुतीकरण को आसान और तुलना को सरल बनाना है। इस सारणी से पाठकों को वांछित सूचना शीघ्र मिल जाती है। इस प्रकार तालिकाएँ विश्लेषक के लिए, कम स्थान में आंकड़ों के विशाल समूह को प्रस्तुत करना संभव बनाती हैं।

आंकड़ों का संग्रह और प्रस्तुतीकरण

आंकड़ों का संग्रह, सारणीयन और सारणी रूप में प्रस्तुतीकरण या तो निरपेक्ष रूप से, प्रतिशत में अथवा संकेत सूची के रूप में होता है।

निरपेक्ष आंकड़ा

जब आंकड़े अपने मूल रूप में पूर्णांक की तरह प्रस्तुत किए जाते हैं, उन्हें निरपेक्ष आंकड़े अथवा कच्चा आंकड़े कहते हैं। उदाहरण के लिए, एक देश अथवा राज्य की कुल जनसंख्या, एक फसल अथवा एक विनिर्माण उद्योग का कुल उत्पादन आदि। *सारणी 1.1* भारत और उसके कुछ चुने हुए राज्यों की जनसंख्या के निरपेक्ष आंकड़े दर्शाती हैं।

प्रतिशत/अनुपात

कई बार आंकड़े अनुपात अथवा प्रतिशत रूप में सारणीबद्ध किए जाते हैं जो कि एक सामान्य प्राचल से परिकलित होते हैं, जैसे साक्षरता दर अथवा जनसंख्या की वृद्धि दर, कृषि उत्पादों अथवा औद्योगिक उत्पादों का प्रतिशत आदि। *सारणी 1.2* विभिन्न दशकों की भारत की साक्षरता दर को प्रतिशत रूप में प्रस्तुत करती

सारणी 1.1 : भारत और चुने हुए राज्यों/केंद्र शासित प्रदेशों की जनसंख्या, 2011

राज्य/केंद्र शासित कोड	भारत/राज्य/केंद्र शासित प्रदेश	कुल जनसंख्या		
		व्यक्ति	पुरुष	स्त्री
1	2	3	4	5
	भारत [1]	**1,21,05,69,573**	**62,31,21,843**	**58,74,47,730**
1.	जम्मू और कश्मीर [2]	1,25,41,302	66,40,662	59,00,640
2.	हिमाचल प्रदेश	68,64,602	34,81,873	33,82,729
3.	पंजाब	2,77,43,338	1,46,39,465	1,31,03,873
4.	चंडीगढ़ [3]	10,55,450	5,80,663	4,74,787
5.	उत्तराखंड	1,00,86,292	51,37,773	49,48,519
6.	हरियाणा	2,53,51,462	1,34,94,734	1,18,56,728
7.	राष्ट्रीय राजधानी प्रदेश, दिल्ली	1,67,87,941	89,87,326	78,00,615
8.	राजस्थान	6,85,48,437	3,55,50,997	3,29,97,440
9.	उत्तर प्रदेश	19,98,12,341	10,44,80,510	9,53,31,831
10.	बिहार	10,40,99,452	5,42,78,157	4,98,21,295

[1] भारत की सभी प्रादेशिक सीमाओं रहित
[2] PoK रहित
[3] केंद्र शासित प्रदेश

स्रोत: 2011 की जनगणना के आंकड़े।

है। साक्षरता दर का परिकलन इस प्रकार किया गया है –

$$\frac{\text{कुल साक्षर व्यक्ति}}{\text{कुल जनसंख्या}} \times 100$$

सारणी 1.2 : साक्षरता दर* : 1951 – 2011

वर्ष	*व्यक्ति*	*पुरुष*	*स्त्री*
1951	18.33	27.16	8.86
1961	28.3	40.4	15.35
1971	34.45	45.96	21.97
1981	43.57	56.38	29.76
1991	52.21	64.13	39.29
2001	64.84	75.85	54.16
2011	73.04	80.9	64.6

स्रोत – 2011 की जनगणना के आंकड़े।

सूचकांक

सूचकांक चर अथवा एक सांख्यिकीय माप है जिसे चर अथवा समय भौगोलिक स्थिति या दूसरी विशेषताओं के संदर्भ में संबंधित चरों के संबंधित समूह में परिवर्तन को दर्शाने के लिए अभिकल्पित किया जाता है। यहाँ ध्यान देने योग्य बात यह है कि सूचकांक न केवल समय के साथ हुए परिवर्तनों की माप करता है बल्कि विभिन्न स्थानों, उद्योगों, नगरों अथवा देशों की आर्थिक दशाओं की तुलना भी करता है। सूचकांक का उपयोग व्यापक रूप में अर्थशास्त्र और व्यवसाय में लागत और मात्रा में आए परिवर्तनों को देखने के लिए किया जाता है। सूचकांक के परिकलन के लिए विभिन्न प्रकार की विधियाँ हैं। फिर भी साधारण समुच्चय विधि सबसे अधिक उपयोग में लाई जाती है। इसे निम्नलिखित सूत्र द्वारा प्राप्त किया जाता है :

$$\frac{q_1}{q_0} \times 100$$

$\sum q_1$ = वर्तमान वर्ष के उत्पादन का योग

$\sum q_0$ = आधार वर्ष के उत्पादन का योग

साधारणतया आधार वर्ष का मूल्य 100 लिया जाता है और उसके आधार पर सूचकांक की गणना की जाती है। उदाहरण के लिए, *तालिका 1.3* भारत में लौह अयस्क के उत्पादन और 1970-71 को आधारवर्ष मानते हुए 1970-71 से 2000-01 तक के सूचकांकों में परिवर्तन को दर्शाती है।

सारणी 1.3 : भारत में लौह अयस्क का उत्पादन

वर्ष	*उत्पादन (मिलियन टन में)*	*परिकलन*	*सूचकांक*
1970-71	32.5	$\frac{32.5}{32.5} \times 100$	100
1980-81	42.2	$\frac{42.2}{32.5} \times 100$	130
1990-91	53.7	$\frac{53.7}{32.5} \times 100$	165
2000-01	67.4	$\frac{67.4}{32.5} \times 100$	207

स्रोत– भारत : आर्थिक सर्वेक्षण, 2005

आंकड़ों का प्रकमण

कच्चे आंकड़ों का प्रक्रमण करने के लिए चयनित वर्गों में उनके सारणीयन और वर्गीकरण की आवश्यकता होती है। उदाहरण के लिए, *तालिका 1.4* में दिए गए आँकड़े का उपयोग यह समझने के लिए किया जा सकता है कि वे किस प्रकार प्रक्रमित किए गए हैं।

हम देख सकते हैं कि दिया गया आंकड़ा अवर्गीकृत हैं। इसलिए सबसे पहला कदम अपरिष्कृत आंकड़ों की मात्रा का बोधगम्य लघुकरण करके उनको वर्गीकृत करना है।

आंकड़ों का वर्गीकरण

कच्चे आंकड़ों के वर्गीकरण के लिए श्रेणियों की संख्याओं को निर्धारित करना होता है जिसमें अपरिष्कृत आंकड़े अपने अंतराल के साथ वर्गीकृत किए जाते हैं। वर्ग अंतराल का चुनाव और वर्गों की संख्या, अपरिष्कृत आंकड़ों के परिसर और वर्गीकरण के उद्देश्यों पर निर्भर करते हैं। तालिका 1.4 में दिए गए कच्चे आंकड़े 2 से 96 तक हैं। सुविधा के लिए हम आंकड़ों को प्रत्येक वर्ग में 10 इकाइयों के अंतराल के साथ, दस वर्गों में रख सकते हैं, उदाहरण के लिए 0–10, 10–20, 20–30 आदि *(तालिका 1.5)*।

सारणी 1.4 : भूगोल विषय में 60 विद्यार्थियों के प्राप्तांक

47	02	39	64	22	46	28	02	09	10
89	96	74	06	26	15	92	84	84	90
32	22	53	62	73	57	37	44	67	50
18	51	36	58	28	65	63	59	75	70
56	58	43	74	64	12	35	42	68	80
64	37	17	31	41	71	56	83	59	90

वर्गीकरण की प्रक्रिया

जब एक बार वर्गों की संख्या और प्रत्येक वर्ग का वर्ग अंतराल निश्चित कर लिया जाता है, तब कच्चे आंकड़ों को वर्गीकृत किया जाता है जैसा कि तालिका 1.5 में दर्शाया गया है। यह एक प्रचलित विधि है जिसे फोर एंड क्रास विधि या मिलान चिह्न के नाम से जाना जाता है।

सबसे पहले, वर्ग की प्रत्येक इकाई के लिए जिसके अंतर्गत वह आता है, एक मिलान चिह्न निर्धारित करते हैं। उदाहरण के लिए, कच्चे आंकड़ों में पहली संख्या 47 है, जो 40-50 के वर्ग में आती है, *सारणी 1.5* के तीसरे कॉलम में एक मिलान चिह्न अंकित कर दिया जाता है।

सारणी 1.5 : आवृत्ति प्राप्त करने के लिए बनाए गए मिलान चिह्न

वर्ग	*कच्चे आंकड़े की संख्या*	*मिलान चिह्न*	*व्यक्ति की संख्या*
0-10	02,02,09,06	////	4
10-20	10,15,18,12,17	~~////~~	5
20-30	22,28,26,22,28	~~////~~	5
30-40	39,32,37,36,35,37,31	~~////~~ //	7
40-50	47,46,44,43,42,41	~~////~~ /	6
50-60	53,57,50,51,58, 59,56,58,56,59	~~////~~ ///	10
60-70	64,62,67,65, 63,64,68,64	~~////~~ ~~////~~	8
70-80	74,73,75,70,74,71	~~////~~ ///	6
80-90	89,84,84,80,83	~~////~~ /	5
90-100	96,92,90,90	~~////~~	4
			$\sum f = N = 60$

आवृत्ति वितरण

तालिका 1.5 में हम मात्रात्मक चरों के कच्चे आंकड़े को वर्गीकृत और उन्हें वर्गानुसार सामूहिक कर चुके हैं। मदों की संख्याएँ (*तालिका 1.5* के चतुर्थ कॉलम में दिए गए स्थान) आवृत्ति कहलाती है और कॉलम आवृत्ति वितरण को प्रदर्शित करता है। यह स्पष्ट होता है कि एक चर की विभिन्न मदों को कैसे वितरित किया गया है। आवृत्तियों को **साधारण** और **संचयी** आवृत्तियों में वर्गीकृत किया जाता है।

सारणी 1.6 : आवृत्ति वितरण

वर्ग	f	Cf
00-10	4	4
10-20	5	9
20-30	5	14
30-40	7	21
40-50	6	27
50-60	10	37
60-70	8	45
70-80	6	51
80-90	5	56
90-100	4	60
	$\sum f = N = 60$	

साधारण आवृत्ति

'f' द्वारा प्रदर्शित साधारण आवृत्ति, प्रत्येक वर्ग के व्यक्तियों की संख्या को प्रदर्शित करती है। (*तालिका 1.6*) सभी वर्गों के लिए दी गई आवृत्ति का योग, दी गई श्रेणी में व्यक्तिगत अवलोकनों के कुल योग को दर्शाता है। सांख्यिकी में, यह 'N' संकेत से स्पष्ट किया गया है जो कि $\sum f$. के बराबर है। इसे $\sum f = N =$ 60 (*तालिका 1.5* और *1.6*) की तरह व्यक्त किया गया है।

संचयी आवृत्ति

संचयी आवृत्ति को '**Cf**', द्वारा प्रदर्शित किया गया है जिसे प्रत्येक वर्ग में दी गई क्रमिक सामान्य आवृत्ति को पहले योग के साथ जोड़कर प्राप्त किया जा सकता है, जैसा कि *तालिका 1.6* के कॉलम 3 में प्रदर्शित है। उदाहरण के लिए *तालिका 1.6* में पहली सामान्य आवृत्ति 4 है। अगली आवृत्ति 5 को 4 में जोड़ा गया है जिसका योग 9 है जो अगली संचयी आवृत्ति है। इसी प्रकार प्रत्येक अगली संख्या को जोड़ते जाते हैं जब तक कि अंतिम संचयी आवृत्ति 60 प्राप्त नहीं हो जाती है। ध्यान देने योग्य बात यह है कि यह N अथवा $\sum f$ के बराबर है।

संचयी आवृत्ति का लाभ यह है कि एक व्यक्ति आसानी से समझ सकता है कि 27 व्यक्ति ऐसे हैं जिनके प्राप्तांक 50 से नीचे हैं अथवा 60 व्यक्तियों में से 45 व्यक्तियों के प्राप्तांक 70 से नीचे हैं।

प्रत्येक सामान्य आवृत्ति इसके समूह अथवा वर्ग से संबंधित होती है। समूहों या वर्गों को तैयार करने के लिए **अपवर्ती** अथवा **समावेशी** विधि प्रयोग में लाई जाती है।

अपवर्ती विधि

जैसा कि *तालिका 1.6* में सबसे पहले कॉलम में दो संख्याएँ दर्शाई गई हैं। ध्यान दें कि एक वर्ग की उच्च सीमा अगले वर्ग की निम्न सीमा के जैसी है। उदाहरण के लिए एक वर्ग (20-30) की उच्च सीमा 30 है जो कि अगले वर्ग (30-40) की निम्न सीमा है। 30 दोनों वर्ग में प्रदर्शित हैं। लेकिन कोई भी अवलोकन जिसका मूल्य 30 है, उसी वर्ग में रखा जाएगा जिसमें यह निम्न सीमा पर आता है और यह उस वर्ग से निकाल दिया जाता है जिसमें यह उच्च सीमा (20-30) पर है। इसीलिए इस विधि को अपवर्ती विधि कहते हैं। अब आप जान सकते हो कि *तालिका 1.4* के सभी सीमांत मूल्य कहाँ जाएँगे।

फिर से *तालिका 1.6* में देखिए, इसके वर्गों की निम्नलिखित प्रकार से व्याख्या की गई है –

0 और 10 से नीचे	10 और 20 से नीचे
20 और 30 से नीचे	30 और 40 से नीचे
40 और 50 से नीचे	50 और 60 से नीचे
60 और 70 से नीचे	70 और 80 से नीचे
80 और 90 से नीचे	90 और100 से नीचे

इस तरह के समूहीकरण में, श्रेणी का विस्तार 10 इकाइयों तक होता है। उदाहरण के लिए 20, 21, 22, 23, 24, 25, 26, 27, 28 और 29 तक संख्याएँ तीसरे वर्ग में आती हैं।

समावेशी विधि

इस विधि में एक मूल्य जो वर्ग की उच्च सीमा के मूल्य के समान होता है, उसे उसी वर्ग में रखा जाता है। इसीलिए इस विधि को समावेशी विधि कहते हैं। इस विधि में वर्गों को अलग प्रकार से प्रदर्शित किया जाता है जैसा *तालिका 1.7* के पहले कॉलम में दिखाया गया है। साधारणतया वर्ग की उच्च सीमा में अगले वर्ग की निम्न सीमा से 1 का अंतर होता है। महत्वपूर्ण बात यह है कि इस विधि में भी वर्ग का विस्तार 10 इकाइयों तक होता है। उदाहरण के लिए 50–59 का वर्ग 10 मानों 50, 51, 52, 53, 54, 55, 56, 57, 58 और 59 (*तालिका 1.7*) का समावेश करता है। इस विधि में उच्च और निम्न दोनों सीमाएँ आवृत्ति वितरण को प्राप्त करने के लिए समाविष्ट की जाती हैं।

सारणी 1.6 : आवृत्ति वितरण : समावेशी विधि

वर्ग	f	Cf
0 – 9	4	4
10 – 19	5	9
20 – 29	5	14
30 – 39	7	21
40 – 49	6	27
50 – 59	10	37
60 – 69	8	45
70 – 79	6	51
80 – 89	5	56
90 – 99	4	60
	$\sum f = N = 60$	

आवृत्ति बहुभुज

आवृत्तियों वितरण का ग्राफ़ आवृत्ति बहुभुज के नाम से जाना जाता है। यह दो या दो से अधिक आवृत्ति वितरण की तुलना में सहायता करता है। दो आवृत्ति को दंड आरेख और रेखाचित्र के द्वारा दिखाया गया है।

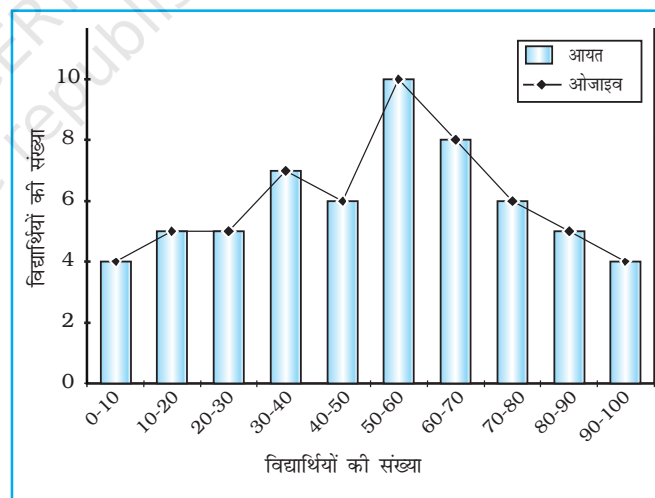
ओजाइव

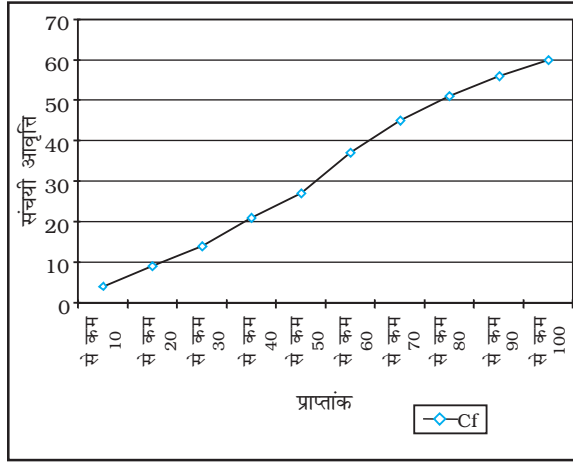
जब आवृत्ति को जोड़ दिया जाता है, उन्हें संचयी आवृत्ति कहा जाता है और जिस सारणी में सूचीगत किए जाते हैं, उसे संचयी आवृत्ति सारणी कहते हैं। संचयी आवृत्ति द्वारा प्राप्त किए गए वक्र को ओजाइव कहते हैं। जिसका उच्चारण ओजाइव है। इसका निर्माण या तो *कमतर विधि* (less than method) या *अधिकतर विधि* (more than method) द्वारा करते हैं।

चित्र 1.5 : *आवृत्ति वितरण बहुभुज*

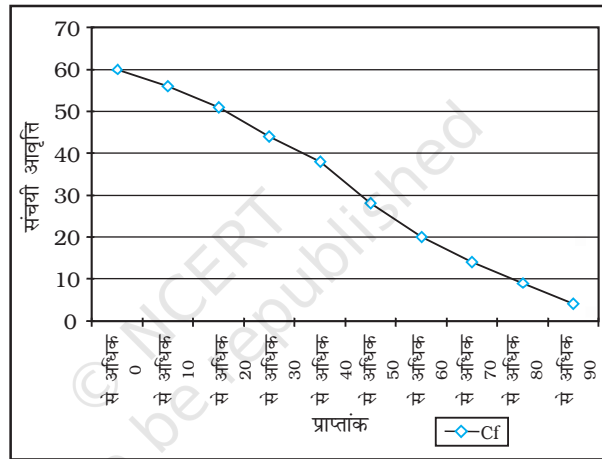
कमतर विधि में, हम श्रेणियों की उच्च सीमा से शुरू करते हैं और आवृत्ति को जोड़ते जाते हैं। जब इन आवृत्तियों को अंकित किया जाता है, तो हमें एक उभरता हुआ वक्र प्राप्त होता है जिसे *तालिका 1.8* और *चित्र 1.5* में दर्शाया गया है।

अधिकतर विधि में, हम वर्गों की निम्न सीमा से शुरू करते हैं और संचयी आवृत्ति से प्रत्येक वर्ग की आवृत्ति को घटा देते हैं। जब ये आवृत्तियाँ अंकित की जाती हैं तब हमें एक गिरता हुआ वक्र प्राप्त होता है जैसा कि *तालिका 1.9* और *चित्र 1.6* में दर्शाया गया है।

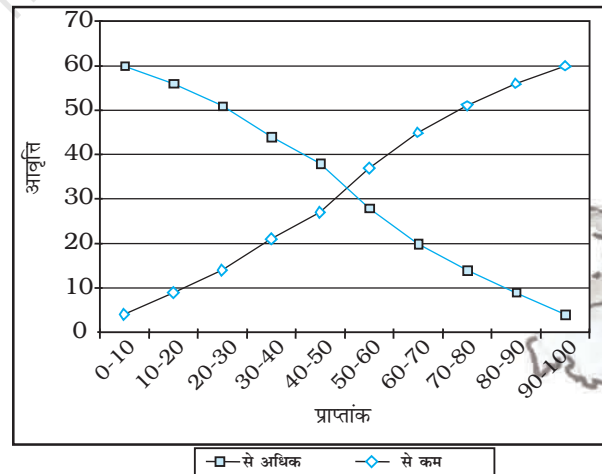
कमतर ओजाइव और अधिकतर ओजाइव का तुलनात्मक चित्र प्राप्त करने के लिए ऊपर के दोनों *चित्रों 1.5* और *1.6* का संयोजन कर सकते हैं जैसा कि *तालिका 1.10* और *चित्र 1.7* में दिखाया गया है।

सारणी 1.8 : आवृत्ति वितरण कमतर विधि

कमतर विधि	Cf
10 से कम	4
20 से कम	9
30 से कम	14
40 से कम	21
50 से कम	27
60 से कम	37
70 से कम	45
80 से कम	51
90 से कम	56
100 से कम	60

चित्र 1.6 : कमतर ओजाइव

सारणी 1.9 : आवृत्ति वितरण अधिकतर विधि

अधिकतर विधि	Cf
0 से अधिक	60
10 से अधिक	56
20 से अधिक	51
30 से अधिक	44
40 से अधिक	38
50 से अधिक	28
60 से अधिक	20
70 से अधिक	14
80 से अधिक	9
90 से अधिक	4

चित्र 1.7: अधिकतर ओजाइव

सारणी 1.10 : कमतर और अधिकतर ओजाइव

प्राप्त प्राप्तांक	कमतर	अधिकतर
0 - 10	4	60
10 - 20	9	56
20 - 30	14	51
30 - 40	21	44
30 - 40	27	38
50 - 60	37	28
60 - 70	45	20
70 - 80	51	14
80 - 90	56	9
90 - 100	60	4

चित्र 1.8 : कमतर और अधिकतर ओजाइव

अभ्यास

1. नीचे दिए गए चार विकल्पों में से सही उत्तर चुनिए :

   (i) एक संख्या अथवा लक्षण को जो मापन को प्रदर्शित करता है, कहते हैं

   (क) अंक (ख) आँकड़े

   (ग) संख्या (घ) लक्षण

   (ii) एकल आधार सामग्री एकमात्र माप है

   (क) तालिका (ख) आवृत्ति

   (ग) वास्तविक संसार (घ) सूचना

   (iii) एक मिलान चिह्न में, फोर एंड क्रांसिंग फिफ्थ द्वारा समूहीकरण को कहते हैं

   (क) फोर एंड क्रास विधि (ख) मिलान चिह्न विधि

   (ग) आवृत्ति अंकित विधि (घ) समावेश विधि

   (iv) ओजाइव एक विधि है जिसमें

   (क) साधारण आवृत्ति नापी जाती है।

   (ख) संचयी आवृत्ति नापी जाती है।

   (ग) साधारण आवृत्ति अंकित की जाती है।

   (घ) संचयी आवृत्ति अंकित की जाती है।

   (v) यदि वर्ग के दोनों अंत आवृत्ति समूह में लिए गए हों, इसे कहते हैं

   (क) बहिष्कार विधि (ख) समावेश विधि

   (ग) चिह्न विधि (घ) सांख्यिकीय विधि

2. निम्नलिखित प्रश्नों के उत्तर लगभग 30 शब्दों में दीजिए :

   (i) आंकड़ा और सूचना के बीच अंतर।

   (ii) आंकड़ों से आप क्या समझते हैं?

   (iii) एक तालिका में पाद टिप्पणी से क्या लाभ हैं?

   (iv) आंकड़ों के प्राथमिक स्रोतों से आपका क्या तात्पर्य है?

   (v) द्वितीयक आंकड़ों के पाँच स्रोत बताइए।

   (vi) आवृत्ति वर्गीकरण की अपवर्ती विधि क्या है?

3. निम्नलिखित प्रश्नों के उत्तर लगभग 125 शब्दों में दीजिए :

   (i) राष्ट्रीय और अंतर्राष्ट्रीय अभिकरणों की चर्चा कीजिए जहाँ से द्वितीयक आँकड़े एकत्र किए जा सकते हैं।

   (ii) सूचकांक का क्या महत्त्व है? सूचकांक की परिकलन की प्रक्रिया को बताने के लिए एक उदाहरण लीजिए और परिवर्तनों को दिखाइए।

क्रियाकलाप

1. भूगोल की 35 विद्यार्थियों की कक्षा में, निम्नलिखित अंक, 10 अंक के यूनिट टेस्ट में प्राप्त किए गए हैं – 1, 0, 2, 3, 4, 5, 6, 7, 2, 3, 4, 0, 2, 5, 8, 4, 5, 3, 6, 3, 2, 7, 6, 5, 4, 3, 7, 8, 9, 7, 9, 4, 5, 4, 3 आँकड़े को संचयी आवृत्ति वितरण के रूप में प्रस्तुत करिए।

   अपनी कक्षा के भूगोल विषय की अंतिम परीक्षा का परिणाम एकत्र कीजिए और प्राप्तांकों को संचयी आवृत्ति वितरण के रूप में प्रदर्शित कीजिए।